

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 16:46:52 ; Search time 144 Seconds
(without alignments)
63.940 Million cell updates/sec

Title: US-10-783-095-10

Perfect score: 131

Sequence: 1 GGGFGAGSLQPLALEGSLQKRGIVEQ 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	100.0	54	US-09-815-229-13	Sequence 13, Appl
2	131	100.0	54	US-10-740-098-13	Sequence 13, Appl
3	131	100.0	86	US-09-878-380-1	Sequence 1, Appl
4	131	100.0	86	US-09-858-935B-4	Sequence 4, Appl
5	131	100.0	86	US-10-028-410-2	Sequence 2, Appl
6	131	100.0	86	US-10-054-873-4	Sequence 4, Appl
7	131	100.0	86	US-10-444-326-2	Sequence 2, Appl
8	131	100.0	86	US-10-271-869-4	Sequence 4, Appl
9	131	100.0	86	US-10-444-262-2	Sequence 2, Appl
10	131	100.0	86	US-10-444-649-2	Sequence 2, Appl
11	131	100.0	86	US-10-444-701-2	Sequence 2, Appl
12	131	100.0	96	US-09-947-563-4	Sequence 4, Appl
13	131	100.0	96	US-09-947-563-5	Sequence 5, Appl

14	131	100.0	110	9	US-09-205-658-125	Sequence 125, Appl
15	131	100.0	110	9	US-09-815-229-3	Sequence 3, Appl
16	131	100.0	110	9	US-09-804-409A-9	Sequence 9, Appl
17	131	100.0	110	10	US-09-969-748C-6	Sequence 6, Appl
18	131	100.0	110	10	US-09-963-893-125	Sequence 125, Appl
19	131	100.0	110	14	US-10-038-886-1	Sequence 1, Appl
20	131	100.0	110	14	US-10-328-813-2	Sequence 2, Appl
21	131	100.0	110	14	US-10-383-285-2	Sequence 2, Appl
22	131	100.0	110	14	US-10-346-563-2	Sequence 2, Appl
23	131	100.0	110	15	US-10-321-717-2	Sequence 2, Appl
24	131	100.0	110	15	US-10-411-037-44	Sequence 44, Appl
25	131	100.0	110	15	US-10-411-026-44	Sequence 44, Appl
26	131	100.0	110	15	US-10-410-962-44	Sequence 44, Appl
27	131	100.0	110	15	US-10-411-049-44	Sequence 44, Appl
28	131	100.0	110	15	US-10-700-725-20	Sequence 20, Appl
29	131	100.0	110	16	US-10-410-930-44	Sequence 44, Appl
30	131	100.0	110	16	US-10-410-997-44	Sequence 44, Appl
31	131	100.0	110	16	US-10-411-012-44	Sequence 44, Appl
32	131	100.0	110	16	US-10-287-994-44	Sequence 44, Appl
33	131	100.0	110	16	US-10-740-098-3	Sequence 3, Appl
34	131	100.0	110	16	US-10-410-913-44	Sequence 44, Appl
35	131	100.0	117	9	US-09-280-030-63	Sequence 63, Appl
36	131	100.0	130	9	US-09-280-030-62	Sequence 62, Appl
37	128	97.7	100	17	US-10-425-115-337310	Sequence 337310,
38	125	95.4	110	16	US-10-419-539-5	Sequence 5, Appl
39	125	95.4	124	15	US-10-221-677-24	Sequence 24, Appl
40	107	81.7	33	17	US-10-641-834-24	Sequence 24, Appl
41	107	81.7	35	9	US-09-815-229-4	Sequence 4, Appl
42	107	81.7	35	9	US-09-947-563-3	Sequence 3, Appl
43	107	81.7	35	16	US-10-740-098-4	Sequence 4, Appl
44	100	76.3	33	9	US-09-815-229-9	Sequence 9, Appl
45	100	76.3	33	16	US-10-740-098-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-815-229-13
; Sequence 13, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 13
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171,147.
US-09-815-229-13

Query Match 100.0%; Score 131; Length 54;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGFGAGSLQPLALEGSLQKRGIVEQ 26

Db 13 GGGFGAGSLQPLALEGSLQKRGIVEQ 38

RESULT 2

US-10-740-098-13
; Sequence 13, Application US/10740098
; Publication No. US20040136101A1
; GENERAL INFORMATION:

```

; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGE DISORDERS
; FILE REFERENCE: P1796RLUS
; CURRENT APPLICATION NUMBER: US/10/740,098
; PRIOR FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 13
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171,147.
US-10-740-098-13

```

```

Query Match          100.0%; Score 131; DB 16; Length 54;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
    |||||
Db 13 GGGPGAGSLQPLALEGSLQKRGIVEQ 38
    |||||

```

```

RESULT 3
US-09-878-380-1
; Sequence 1, Application US/09878380
; Patent No. US20020160435A1
; GENERAL INFORMATION:
; APPLICANT: Fujirebio Inc.
; APPLICANT: KITAJIMA, Sachiko
; APPLICANT: KURANO, Yoshihiro
; APPLICANT: NAKATSUBO, Kaoru
; APPLICANT: NISHIZONO, Isao
; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
; FILE REFERENCE: 0760-0291P
; CURRENT APPLICATION NUMBER: US/09/878,380
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: JP 2000-174691
; PRIOR FILING DATE: 2000-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-878-380-1

```

```

Query Match          100.0%; Score 131; DB 9; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
    |||||
Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70
    |||||

```

```

RESULT 4
US-09-858-935B-4
; Sequence 4, Application US/09858935B
; Publication No. US20030069177A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02

```

```

; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 4
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-935B-4

```

```

Query Match          100.0%; Score 131; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
    |||||
Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70
    |||||

```

```

RESULT 5
US-10-028-410-2
; Sequence 2, Application US/10028410
; Publication No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-2

```

```

Query Match          100.0%; Score 131; DB 13; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
    |||||
Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70
    |||||

```

```

RESULT 6
US-10-054-873-4
; Sequence 4, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intra-molecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 22, 2004, 20:14:29 ; Search time 339 Seconds
(without alignments)
414.437 Million cell updates/sec

Title: US-10-783-095-10

Perfect score: 131

Sequence: 1 GGGPGAGSLQALEGSLQKRGIVEQ_26

Scoring table:
Xgapop 10.0, Xgapext 0.5
Igapop 10.0, Igapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model
-Q=/cgn2_1/USPTO_spool/US10783095/runat_22112004_143538_25412/app_query.fasta_1.199
-DB=Published Applications NA -CEMT=fastap -SUFFIX=P2N.rnpb -MINMATCH=0.1
-LOOPCL=0 -L_OOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095@cgn_1.1.575@runat_22112004_143538_25412
-NCPV=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	131	100.0	258	15	US-10-383-285-1	Sequence 1, Appl
2	131	100.0	260	10	US-09-807-742-16	Sequence 16, Appl
3	131	100.0	260	10	US-09-807-742-17	Sequence 17, Appl
4	131	100.0	260	13	US-10-746-149-23	Sequence 23, Appl
5	131	100.0	260	18	US-10-609-019-21	Sequence 21, Appl
6	131	100.0	330	15	US-10-346-563-1	Sequence 1, Appl
7	131	100.0	330	16	US-10-321-717-1	Sequence 1, Appl
8	131	100.0	351	9	US-09-280-030-49	Sequence 49, Appl
9	131	100.0	390	9	US-09-280-030-48	Sequence 48, Appl
10	131	100.0	450	10	US-09-804-409A-10	Sequence 10, Appl
11	131	100.0	450	10	US-09-969-748C-5	Sequence 5, Appl
12	131	100.0	450	15	US-10-328-813-1	Sequence 1, Appl
13	131	100.0	450	16	US-10-411-037-43	Sequence 43, Appl
14	131	100.0	450	16	US-10-411-026-43	Sequence 43, Appl
15	131	100.0	450	16	US-10-410-962-43	Sequence 43, Appl
16	131	100.0	450	16	US-10-411-049-43	Sequence 43, Appl
17	131	100.0	450	17	US-10-410-930-43	Sequence 43, Appl
18	131	100.0	450	17	US-10-410-997-43	Sequence 43, Appl
19	131	100.0	450	17	US-10-411-012-43	Sequence 43, Appl
20	131	100.0	450	17	US-10-287-994-43	Sequence 43, Appl
21	131	100.0	450	17	US-10-410-913-43	Sequence 43, Appl
22	131	100.0	1182	9	US-09-919-344-27	Sequence 27, Appl
23	131	100.0	1626	18	US-10-609-019-27	Sequence 27, Appl
24	131	100.0	2499	10	US-09-775-508A-1	Sequence 1, Appl
25	131	100.0	10487	17	US-10-746-149-44	Sequence 44, Appl
26	131	100.0	10487	18	US-10-609-019-32	Sequence 32, Appl
27	131	100.0	10512	17	US-10-746-149-42	Sequence 42, Appl
28	131	100.0	10512	18	US-10-609-019-31	Sequence 31, Appl
29	131	100.0	10880	17	US-10-746-149-47	Sequence 47, Appl
30	131	100.0	10895	18	US-10-609-019-41	Sequence 41, Appl
31	131	100.0	11255	17	US-10-746-149-43	Sequence 43, Appl
32	131	100.0	11271	18	US-10-609-019-42	Sequence 42, Appl
33	131	100.0	11332	18	US-10-609-019-43	Sequence 43, Appl
34	128	97.7	348	18	US-10-425-115-152647	Sequence 152647,
35	125	95.4	333	17	US-10-419-539-2	Sequence 2, Appl
36	125	95.4	375	16	US-10-221-677-25	Sequence 25, Appl
37	125	95.4	598	13	US-10-013-032-11	Sequence 11, Appl
38	125	95.4	633	13	US-10-013-032-12	Sequence 12, Appl
39	125	95.4	666	13	US-10-013-032-13	Sequence 13, Appl
40	125	95.4	1080	17	US-10-419-539-3	Sequence 3, Appl
41	125	95.4	1217	17	US-10-419-539-4	Sequence 4, Appl
42	125	95.4	1944	16	US-10-221-677-26	Sequence 26, Appl
43	111	84.7	498	9	US-09-925-297-346	Sequence 346, App
44	99	75.6	4646	16	US-10-430-752A-14	Sequence 14, App
45	98	74.8	390	9	US-09-833-381-939	Sequence 939, App

ALIGNMENTS

RESULT 1
US-10-383-285-1 ; Sequence 1, Application US/10383285
; Publication No. US20030200566A1
; GENERAL INFORMATION:
; APPLICANT: STREATHFIELD, STEPHEN
; APPLICANT: HOWARD, JOHN
; TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN PLANTS
; FILE REFERENCE: 10040
; CURRENT APPLICATION NUMBER: US/10/383,285
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: 60/362,874
; PRIOR FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-383-285-1

```

Alignment Scores:
Pred. No.: 4.7e-11 Length: 258
Score: 131.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-10-383-285-1 (1-258)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
Db 133 GCGGTGGCCCGGGCCGGCTCCCTCCAGCGCTCGCCCTCGAGGGTCCCTCCAGAG 192
Qy 21 ArgGlyIleValGluGln 26
Db 193 CCGGTATCGTGGACAG 210

```

```

RESULT 2
US-10-746-149-23
; Sequence 23, Application US/09807742
; Publication No. US20030204864A1
; GENERAL INFORMATION:
; APPLICANT: DANIELL, HENRY
; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
; FILE REFERENCE: 1465-PCT-US-00
; CURRENT APPLICATION NUMBER: US/09/807,742
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/06288
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-807-742-16

```

```

Alignment Scores:
Pred. No.: 4.73e-11 Length: 260
Score: 131.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-09-807-742-16 (1-260)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
Db 133 GCGGGGCGCCCTGGTGGCAGCGCTCGACCCCTTGGCCCTGGAGGGTCCCTCCAGAG 192
Qy 21 ArgGlyIleValGluGln 26
Db 193 CCGTGCATGTGGACAA 210

```

```

RESULT 3
US-09-807-742-17
; Sequence 17, Application US/09807742
; Publication No. US20030204864A1
; GENERAL INFORMATION:
; APPLICANT: DANIELL, HENRY
; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
; FILE REFERENCE: 1465-PCT-US-00
; CURRENT APPLICATION NUMBER: US/09/807,742
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/06288
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 17
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
; OTHER INFORMATION: modified proinsulin sequence
US-09-807-742-17

```

```

Alignment Scores:
Pred. No.: 4.73e-11 Length: 260
Score: 131.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-09-807-742-17 (1-260)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
Db 133 GCGGGGCGCCCTGGTGGCAGCGCTCGACCCCTTGGCCCTTACAGTTAGAGGTTCTTTACAAA 192
Qy 21 ArgGlyIleValGluGln 26
Db 193 CCGTGCATGTGGACAA 210

```

```

RESULT 4
US-10-746-149-23
; Sequence 23, Application US/10746149
; Publication No. US20040172667A1
; GENERAL INFORMATION:
; APPLICANT: Cooper, Richard K.
; APPLICANT: Fioretti, William C.
; APPLICANT: Cadd, Gary G.
; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
; FILE REFERENCE: S1687-0280 (S1687-294923)
; CURRENT APPLICATION NUMBER: US/10/746,149
; CURRENT FILING DATE: 2003-12-24
; PRIOR APPLICATION NUMBER: US 60/392,415
; PRIOR FILING DATE: 2002-06-26
; PRIOR APPLICATION NUMBER: US 60/441,381
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,447
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,405
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,502
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,392
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: US 60/441,377
; PRIOR APPLICATION NUMBER: US 10/609,019
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-746-149-23

```

```

Alignment Scores:
Pred. No.: 4.73e-11 Length: 260
Score: 131.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-09-807-742-17 (1-260)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
Db 133 GCGGGGCGCCCTGGTGGCAGCGCTCGACCCCTTGGCCCTTACAGTTAGAGGTTCTTTACAAA 192
Qy 21 ArgGlyIleValGluGln 26
Db 193 CCGTGCATGTGGACAA 210

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 16:59:23 ; Search time 37 Seconds

(without alignments)

46.602 Million cell updates/sec

Title: US-10-783-095-10

Perfect score: 26

Sequence: 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance by a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length DB, ID, Description. Contains 27 rows of search results.

Table with columns: Seq ID, Score, Query Match, Length DB, ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1
US-09-815-229-13
; Sequence 13, Application US/09815229
; Patent No. 6689747
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 13
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171,147.
US-09-815-229-13

Query Match 100.0%; Score 26; DB 4; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.8e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
| | | | |
Db 13 GGGPGAGSLQPLALEGSLQKRGIVEQ 38

RESULT 2
US-09-477-924-2
; Sequence 2, Application US/09477924
; Patent No. 6403764
; GENERAL INFORMATION:
; APPLICANT: Dubague, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/09/477,924
; CURRENT FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-477-924-2

Query Match 100.0%; Score 26; DB 4; Length 86;

Best Local Similarity 100.0%; Pred. No. 5.7e-18; Indels 0; Gaps 0; Mismatches 0; Conservative 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
DB 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 3

US-09-723-981-2
Sequence 2, Application US/097233981
Patent No. 6506874
GENERAL INFORMATION:
APPLICANT: Dubaqui, Yves
APPLICANT: Lowman, Henry
TITLE OF INVENTION: PROTEIN VARIANTS
FILE REFERENCE: P1712R1
CURRENT APPLICATION NUMBER: US/09/723,981
CURRENT FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 09/477,923
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 6
SEQ ID NO 2
LENGTH: 86
TYPE: PRT
ORGANISM: Homo sapiens
US-09-723-981-2

Query Match 100.0%; Score 26; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 5.7e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
DB 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 4

US-09-723-896-2
Sequence 2, Application US/097233896
Patent No. 6509443
GENERAL INFORMATION:
APPLICANT: Dubaqui, Yves
APPLICANT: Lowman, Henry
TITLE OF INVENTION: PROTEIN VARIANTS
FILE REFERENCE: P1712R1
CURRENT APPLICATION NUMBER: US/09/723,896
CURRENT FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: US/09/477,923
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 6
SEQ ID NO 2
LENGTH: 86
TYPE: PRT
ORGANISM: Homo sapiens
US-09-723-896-2

Query Match 100.0%; Score 26; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 5.7e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
DB 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 5

US-09-878-380-1
Sequence 1, Application US/09878380
Patent No. 6534281
GENERAL INFORMATION:
APPLICANT: Fujirebio Inc.
APPLICANT: KITAJIMA, Sachiko

APPLICANT: KURANO, Yoshihiro
APPLICANT: NAKATSUBO, Kaoru
APPLICANT: NISHIZONO, Isao
TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
FILE REFERENCE: 0760-0291P
CURRENT APPLICATION NUMBER: US/09/878,380
CURRENT FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: JP 2000-174691
PRIOR FILING DATE: 2000-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 86
TYPE: PRT
ORGANISM: Homo sapiens
US-09-878-380-1

Query Match 100.0%; Score 26; DB 4; Length 86;
Best Local Similarity 100.0%; Pred. No. 5.7e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
DB 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 6

US-08-030-731A-43
Sequence 43, Application US/08030731A
Patent No. 5426036
GENERAL INFORMATION:
APPLICANT: Koller, Klaus-Peter
APPLICANT: Riess, Guenther Johannes
APPLICANT: Uhlmann, Eugen
APPLICANT: Wallmeier, Holger
TITLE OF INVENTION: Processes for the Preparation of Foreign
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/030,731A
FILING DATE: 12-MAR-1993
CLASSIFICATION: 435
APPLICATION NUMBER: US 07/189,840
FILING DATE: 03-MAY-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/430,622
FILING DATE: 01-NOV-1989
APPLICATION NUMBER: US 07/687,610
FILING DATE: 19-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/735,757
FILING DATE: 29-JUL-1991
APPLICATION NUMBER: DE P 37 14 866.4
FILING DATE: 05-MAY-1987
APPLICATION NUMBER: DE P 38 37 273.8
FILING DATE: 03-NOV-1988

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 17:07:53 ; Search time 140 Seconds
(without alignments)
65.767 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGGPAGSLQPLALEGSLQKRGIVEQ 26

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1570615 seqs, 354127592 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubppaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubppaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	54	9	US-09-815-229-13
2	26	100.0	54	16	US-10-740-098-13
3	26	100.0	86	9	US-09-878-380-1
4	26	100.0	86	10	US-09-858-935B-4
5	26	100.0	86	13	US-10-028-410-2
6	26	100.0	86	13	US-10-054-873-4
7	26	100.0	86	14	US-10-444-326-2
8	26	100.0	86	14	US-10-271-869-4
9	26	100.0	86	15	US-10-444-262-2
10	26	100.0	86	15	US-10-444-649-2
11	26	100.0	86	15	US-10-444-701-2
12	26	100.0	96	9	US-09-947-563-4
13	26	100.0	96	9	US-09-947-563-5

14	26	100.0	110	9	US-09-205-658-125
15	26	100.0	110	9	US-09-815-229-3
16	26	100.0	110	9	US-09-804-409A-9
17	26	100.0	110	10	US-09-969-748C-6
18	26	100.0	110	14	US-09-963-693-125
19	26	100.0	110	14	US-10-038-686-1
20	26	100.0	110	14	US-10-328-813-2
21	26	100.0	110	14	US-10-383-285-2
22	26	100.0	110	14	US-10-346-563-2
23	26	100.0	110	15	US-10-321-717-2
24	26	100.0	110	15	US-10-411-037-44
25	26	100.0	110	15	US-10-411-026-44
26	26	100.0	110	15	US-10-410-962-44
27	26	100.0	110	15	US-10-411-049-44
28	26	100.0	110	15	US-10-705-725-20
29	26	100.0	110	16	US-10-410-930-44
30	26	100.0	110	16	US-10-410-997-44
31	26	100.0	110	16	US-10-411-012-44
32	26	100.0	110	16	US-10-287-994-44
33	26	100.0	110	16	US-10-740-098-3
34	26	100.0	110	16	US-10-410-913-44
35	26	100.0	117	9	US-09-280-030-63
36	26	100.0	130	9	US-09-280-030-62
37	24	92.3	100	17	US-10-425-115-337310
38	21	80.8	33	17	US-10-641-834-24
39	21	80.8	35	9	US-09-815-229-4
40	21	80.8	35	9	US-09-947-563-3
41	21	80.8	35	16	US-10-740-098-4
42	19	73.1	31	9	US-09-269-438-1
43	19	73.1	31	9	US-09-878-380-2
44	19	73.1	31	15	US-10-430-752A-1
45	19	73.1	31	15	US-10-430-752A-15

ALIGNMENTS

RESULT 1
 US-09-815-229-13
 ; Sequence 13, Application US/09815229
 ; Patent No. US20020058614A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Filvaroff, Ellen H.
 ; APPLICANT: Okumu, Franklin W.
 ; FILE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
 ; TITLE REFERENCE: P1786RIUS
 ; CURRENT APPLICATION NUMBER: US/09/815,229
 ; CURRENT FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: US 60/192,103
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 17
 ; SEQ ID NO 13
 ; LENGTH: 54
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Portion of insulin variant from EP 171,147.
 US-09-815-229-13

Query Match 100.0%; Score 26; DB 9; Length 54;
 Best Local Similarity 100.0%; Pred. No. 2.9e-16;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPAGSLQPLALEGSLQKRGIVEQ 26
 |||
 Db 13 GGGPAGSLQPLALEGSLQKRGIVEQ 38

RESULT 2
 US-10-740-098-13
 ; Sequence 13, Application US/10740098
 ; Publication No. US20040138101A1
 ; GENERAL INFORMATION:

; APPLICANT: Filvaroff, Ellen H.
 ; APPLICANT: Okumu, Franklin W.
 ; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
 ; FILE REFERENCE: P1786R1US
 ; CURRENT APPLICATION NUMBER: US/10/740,098
 ; CURRENT FILING DATE: 2003-12-17
 ; PRIOR APPLICATION NUMBER: US/09/815,229
 ; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: US 60/192,103
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 17
 ; SEQ ID NO 13
 ; LENGTH: 54
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Portion of insulin variant from EP 171,147.
 US-10-740-098-13

Query Match 100.0%; Score 26; DB 16; Length 54;
 Best Local Similarity 100.0%; Pred. No. 2.9e-16;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 Db 13 GGGPGAGSLQPLALEGSLQKRGIVEQ 38

RESULT 3
 ; Sequence 1, Application US/09878380
 ; Patent No. US20020160435A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fujirebio Inc.
 ; APPLICANT: KIRAJIMA, Sachiko
 ; APPLICANT: KURANO, Yoshihiro
 ; APPLICANT: NAKATSUBO, Kaoru
 ; APPLICANT: NISHIZONO, Isao
 ; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
 ; CURRENT APPLICATION NUMBER: US/09/878,380
 ; CURRENT FILING DATE: 2001-06-12
 ; PRIOR APPLICATION NUMBER: JP 2000-174691
 ; PRIOR FILING DATE: 2000-06-12
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 86
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-878-380-1

Query Match 100.0%; Score 26; DB 9; Length 86;
 Best Local Similarity 100.0%; Pred. No. 4.3e-16;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 4
 ; Sequence 4, Application US/09858935B
 ; Publication No. US20030069177A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dubaqui, Yves
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Lowman, Henry B.
 ; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
 ; FILE REFERENCE: P1794RI
 ; CURRENT APPLICATION NUMBER: US/09/858,935B
 ; CURRENT FILING DATE: 2002-07-02

; PRIOR APPLICATION NUMBER: US 60/248,985
 ; PRIOR FILING DATE: 2000-11-15
 ; PRIOR APPLICATION NUMBER: US 60/204,490
 ; PRIOR FILING DATE: 2000-05-16
 ; NUMBER OF SEQ ID NOS: 153
 ; SEQ ID NO 4
 ; LENGTH: 86
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-858-935B-4

Query Match 100.0%; Score 26; DB 10; Length 86;
 Best Local Similarity 100.0%; Pred. No. 4.3e-16;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 5
 ; Sequence 2, Application US/10028410
 ; Publication No. US20020160955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dubaqui, Yves
 ; APPLICANT: Lowman, Henry
 ; TITLE OF INVENTION: PROTEIN VARIANTS
 ; FILE REFERENCE: P1712R1-1
 ; CURRENT APPLICATION NUMBER: US/10/028,410
 ; CURRENT FILING DATE: 2001-12-19
 ; PRIOR APPLICATION NUMBER: US/09/477,924
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 6
 ; SEQ ID NO 2
 ; LENGTH: 86
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-028-410-2

Query Match 100.0%; Score 26; DB 13; Length 86;
 Best Local Similarity 100.0%; Pred. No. 4.3e-16;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 Db 45 GGGPGAGSLQPLALEGSLQKRGIVEQ 70

RESULT 6
 ; Sequence 4, Application US/10054873
 ; Publication No. US20020164712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gan, Zhong Ru
 ; TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/10/054,873
 ; FILING DATE: 22-Jan-2002

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 16:50:52 ; Search time 193 Seconds
(without alignments)
77.512 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	65	Q8HZ80	Q8hz80 pongo pygma
2	26	100.0	65	Q8HZ81	Q8hz81 gorilla gor
3	26	100.0	110	INS_CERAE	P30407 cercopithe
4	26	100.0	110	INS_HUMAN	P01308 homo sapien
5	26	100.0	110	INS_MACFA	P30406 macaca fasc
6	26	100.0	110	INS_PANTR	P30410 pan troglod
7	26	100.0	110	INS_PONPY	Q8hxv2 pongo pygma
8	26	100.0	110	O6YK33	O6yk33 gorilla gor
9	26	100.0	110	AAP35454	Aap35454 homo sapi
10	26	100.0	110	AAN06935	Aan06935 gorilla g
11	26	100.0	110	AAN39451	Aan39451 homo sapi
12	9	34.6	110	INS_CANFA	P01321 canis fami
13	9	34.6	110	INS_RABIT	P01311 cryptolagus
14	9	34.6	110	INS_SPETR	Q8ix13 spermophilu
15	9	34.6	110	Q8WRW6	Q8wrw6 felis silve
16	8	30.8	105	INS_BOVIN	P01317 bos taurus
17	8	30.8	105	INS_SHEEP	P01318 ovis aries
18	8	30.8	108	INS_PIG	P01315 sus scrofa
19	8	30.8	108	AAQ00952	AAq00952 sus scrof
20	8	30.8	108	AAQ00954	AAq00954 sus scrof
21	8	30.8	108	AAQ00957	AAq00957 sus scrof
22	8	30.8	108	AAQ00960	AAq00960 sus scrof
23	8	30.8	108	AAQ00963	AAq00963 sus scrof
24	8	30.8	108	AAQ00966	AAq00966 sus scrof
25	8	30.8	108	AAQ00969	AAq00969 sus scrof
26	8	30.8	108	AAQ00972	AAq00972 sus scrof
27	8	30.8	108	AAQ00975	AAq00975 sus scrof
28	8	30.8	108	AAQ00978	AAq00978 sus scrof
29	8	30.8	108	AAQ00981	AAq00981 sus scrof
30	8	30.8	108	AAQ00983	AAq00983 sus scrof
31	8	30.8	108	AAQ00985	AAq00985 sus scrof

32 8 30.8 108 2 AAQ00987 Aaq00987 sus scrof
 33 8 30.8 108 2 AAQ00990 Aaq00990 sus scrof
 34 8 30.8 110 1 INS_PSAOB Q62587 psammomys o
 35 8 30.8 447 2 Q9SIA8 Q9sia8 arabidopsis
 36 8 30.8 465 1 FXD3_MOUSE Q61060 mus musculu
 37 8 30.8 478 1 FXD3_HUMAN Q9juj5 homo sapien
 38 8 30.8 485 2 Q8PNW5 Q8pnw5 xanthomonas
 39 8 30.8 549 2 Q8Z9A9 Q8z9a9 streptomyce
 40 8 30.8 605 2 Q6DI71 Q6di71 mus musculu
 41 8 30.8 631 1 DMK_MOUSE P54265 mus musculu
 42 7 26.9 12 2 Q6QHB8 Q6qhb8 homo sapien
 43 7 26.9 12 2 AAS5053 Aas5053 homo sapi
 44 7 26.9 38 2 Q6XV13 Q6xv13 strix aluco
 45 7 26.9 38 2 Q6XVJ2 Q6xvj2 meleagris g

ALIGNMENTS

RESULT 1
 Q8HZ80 PRELIMINARY; PRT; 65 AA.
 AC Q8HZ80;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Insulin (Fragment)
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Sutheria; Primates; Catarrhini; Hominidae; Pongo.
 OX NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'hujigin C., Tichy H., Klein J.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY092024; AAM76641.1; -
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological process; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 FT NON_TER 1
 FT NON_TER 65
 SQ SEQUENCE 65 AA; 6920 MW; E772017FDBSECABEA CRC64;
 Query Match 100.0%; Score 26; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 4e-17;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 |||||
 Db 39 GGGPGAGSLQPLALEGSLQKRGIVEQ 64
 |||||

RESULT 2
 Q8HZ81 PRELIMINARY; PRT; 65 AA.
 AC Q8HZ81;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Insulin (Fragment).
 OS Gorilla gorilla (Gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Sutheria; Primates; Catarrhini; Hominidae; Gorilla.
 OX NCBI_TaxID=9593;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'hujigin C., Tichy H., Klein J.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR EMBL; AY092023; AAM76640.1; -

DR GO:0005576; C:extracellular; IEA.
 DR GO:0005179; F:hormone activity; IEA.
 DR GO:0007582; P:physiological process; IEA.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; I.
 DR SMART; SM00078; ILGF; 1.
 KW Insulin family.
 FT NON_TER 1
 FT TER 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 100.0%; Score 26; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 4e-17;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR GO:0005576; C:extracellular; IEA.
 DR GO:0005179; F:hormone activity; IEA.
 DR GO:0007582; P:physiological process; IEA.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; I.
 DR SMART; SM00078; ILGF; 1.
 KW Insulin family.
 FT NON_TER 1
 FT TER 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 100.0%; Score 26; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 4e-17;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 DB 39 GGGPGAGSLQPLALEGSLQKRGIVEQ 64

OY 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 DB 39 GGGPGAGSLQPLALEGSLQKRGIVEQ 64

RESULT 3
 INS CERAE STANDARD; PRT; 110 AA.
 AC P30407; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.
 RT "Sequences of primate insulin genes support the hypothesis of a slower
 rate of molecular evolution in humans and apes than in monkeys."
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=72258016; PubMed=4626369;
 RA Peterson J.D., Mehrlich S., Oyer P.E., Steiner D.F.;
 RT "Determination of the amino acid sequence of the monkey, sheep, and
 dog proinsulin C-peptides by a semi-micro Edman degradation
 procedure."
 RL J. Biol. Chem. 247:4866-4871(1972).
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the insulin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X61092; CAA43405.1; -
 DR PIR; B42179; B42179.
 DR HSSP; P01308; 1A1C.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINE.
 DR SMART; SM00078; ILGF; 1.

RESULT 4
 INS HUMAN STANDARD; PRT; 110 AA.
 AC P01308.
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80120725; PubMed=6243748;
 RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
 RA Goodman H.M.;
 RT "Sequence of the human insulin gene."
 RL Nature 284:26-32(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80236313; PubMed=6248962;
 RA Ullrich A., Dull T.J., Gray A., Sures I.;
 RT "Genetic variation in the human insulin gene."
 RL Science 209:612-615(1980).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80054779; PubMed=503234;
 RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
 RA Rutter W.J.;
 RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin."
 RL Nature 282:525-527(1979).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80147417; PubMed=6927840;
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
 RT "Nucleotide sequence of human preproinsulin complementary DNA."
 RL Science 208:57-59(1980).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93364428; PubMed=8358440;
 RA Lucassen A.M., Bell J.I., Jullier C., Lathrop M.;
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
 kb segment of DNA spanning the insulin gene and associated VNTR."
 RL Nat. Genet. 4:305-310(1993).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Pancreas;
 RL MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feigold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 16:58:57 ; Search time 38 Seconds
(without alignments)
65.833 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGGPGAGSLQPLALEGSLQKRGIVEV 26

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR.791.*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	26	100.0	110	1	IPHU
2	26	100.0	110	2	B42179
3	26	100.0	110	3	JQ0178
4	26	100.0	110	2	A42179
5	9	34.6	110	1	INRB
6	9	34.6	110	1	IPDG
7	8	30.8	84	1	IPFG
8	8	30.8	105	1	IPBO
9	8	30.8	412	2	I78395
10	8	30.8	441	2	I78393
11	8	30.8	447	2	G84687
12	8	30.8	474	2	I78396
13	8	30.8	515	2	I78394
14	8	30.8	557	2	S71829
15	7	26.9	103	2	I51221
16	7	26.9	105	1	IPON
17	7	26.9	106	1	IPXL1
18	7	26.9	106	1	IPXL2
19	7	26.9	107	1	IPCH
20	7	26.9	108	1	IPCA
21	7	26.9	109	1	IPRTDU
22	7	26.9	110	1	INMS2
23	7	26.9	110	1	IPGP
24	7	26.9	110	1	IPRT2
25	7	26.9	115	1	IPAF
26	7	26.9	115	1	IPBF
27	7	26.9	234	2	G95989
28	7	26.9	245	2	A70777
29	7	26.9	277	1	ESPSSK

30	7	26.9	286	1	C35124
31	7	26.9	286	2	JND816
32	7	26.9	297	2	T28709
33	7	26.9	335	2	B90044
34	7	26.9	371	2	I46089
35	7	26.9	380	2	A70646
36	7	26.9	406	2	E70675
37	7	26.9	412	2	H81713
38	7	26.9	431	2	F84077
39	7	26.9	456	2	T40386
40	7	26.9	473	2	G84312
41	7	26.9	482	2	H70463
42	7	26.9	495	1	A26396
43	7	26.9	502	2	D75030
44	7	26.9	502	2	B71216
45	7	26.9	505	2	S39962

ALIGNMENTS

RESULT 1

IPHU
insulin precursor [validated] - human
N;Alternate names: preproinsulin
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1981 #sequence revision 23-Oct-1981 #text change 05-Jul-2004
C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S581
R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischler, E.; Goodman, H.M.
Nature 284, 26-32, 1980
A;Title: Sequence of the human insulin gene.
A;Reference number: A93222; MUID:80120725; PMID:6243748
A;Accession: A93222
A;Molecule type: DNA
A;Residues: 1-110 <BEL>
A;Cross-references: UNIPROT:P01306; GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
Science 209, 612-615, 1980
A;Title: Genetic variation in the human insulin gene.
A;Reference number: A94253; MUID:80236313; PMID:6248962
A;Accession: A94253
A;Molecule type: DNA
A;Residues: 1-110 <ULL>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
Nature 282, 525-527, 1979
A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
A;Reference number: A93216; MUID:80054779; PMID:503234
A;Accession: A93216
A;Molecule type: mRNA
A;Residues: 1-110 <BEL2>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
Science 208, 57-59, 1980
A;Title: Nucleotide sequence of human preproinsulin complementary DNA.
A;Reference number: A94251; MUID:80147417; PMID:6927840
A;Accession: A94251
A;Molecule type: mRNA
A;Residues: 1-110 <SUR>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Nicol, D.S.H.W.; Smith, L.F.
Nature 187, 483-485, 1960
A;Title: Amino-acid sequence of human insulin.
A;Reference number: A93144
A;Accession: A93144
A;Molecule type: protein
A;Residues: 25-54; 90-110 <NIC>
R;Over, P.E.; Cho, S.; Peterson, J.D.; Stelner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan
A;Reference number: A92075; MUID:71116410; PMID:5101771
A;Accession: A92075
A;Molecule type: protein

A;Residues: 57-87 <OYE>
 R;KO, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A;Title: Amino acid sequence of the C-peptide of human proinsulin.
 A;Reference number: A91186; MUID:71257722; PMID:5560404
 A;Accession: A91186
 A;Molecule type: protein
 A;Residues: 57-87 <KOB>
 R;Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
 A;Reference number: I58114; MUID:93364428; PMID:8358440
 A;Accession: I58114
 A;Status: Preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-59,63-110 <RES>
 A;Cross-references: GB:L15440; NID:G307071; PIDN:AAA59179.1; PID:G307072
 R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A;Reference number: A91636; MUID:7507277; PMID:4443293
 A;Contents: annotation; synthesis
 A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
 A;Note: article in German with English abstract
 R;Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A;Title: The synthesis of C-peptide of human proinsulin.
 A;Reference number: A91658; MUID:75040007; PMID:4803504
 A;Contents: annotation; synthesis of residues 57-87
 R;Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
 A;Reference number: A90914
 A;Contents: annotation; synthesis of residues 57-87
 R;Kaufman, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide juncti
 A;Reference number: S58661; MUID:96013185; PMID:7575420
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C;Genetics:
 A;Gene: GDB:INS
 A;Cross-references: GDB:119349; OMIM:176730
 A;Map position: 11p15.5-11p15.5
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status experimental <BCH>
 F;55-87/Domain: connecting C peptide #status experimental <CEPP>
 F;90-110/Domain: insulin chain A #status experimental <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 100.0%; Score 26; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred.No. 4.4e-18;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 |||
 Db 69 GGGPGAGSLQPLALEGSLQKRGIVEQ 94
 RESULT 2
 A42179
 Insulin precursor - green monkey
 C;Species: Cercopithecus aethiops (green monkey, grivet)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: B42179; A05232; S16494; S22056
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219553; PMID:1560757
 A;Accession: B42179

A;Molecule type: DNA
 A;Residues: 1-110 <SEI>
 A;Cross-references: UNIPROT:P30407; EMBL:X61092; NID:G22808; PIDN:CAA43405.1; PID:G22809
 A;Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIF:95194)
 R;Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin;
 A;Reference number: A92111; MUID:72258016; PMID:4626369
 A;Accession: A05232
 A;Molecule type: protein
 A;Residues: 57-87 <PET>
 C;Genetics:
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status predicted <BCH>
 F;55-87/Domain: connecting peptide #status experimental <CEPP>
 F;90-110/Domain: insulin chain A #status predicted <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 26; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred.No. 4.4e-18;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 |||
 Db 69 GGGPGAGSLQPLALEGSLQKRGIVEQ 94
 RESULT 3
 JQ0178
 Insulin precursor - crab-eating macaque
 C;Species: Macaca fascicularis (crab-eating macaque)
 C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 R;Katsamian, J.Q0178
 C;Accession: JQ0178
 R;Wetkamm, W.; Gronberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
 Gene 19, 179-183, 1982
 A;Title: The nucleotide sequence of cDNA coding for proinsulin from the primate Macaca
 A;Reference number: JQ0178; MUID:83080474; PMID:6184262
 A;Molecule type: mRNA
 A;Accession: JQ0178
 A;Residues: 1-110 <WET>
 A;Cross-references: UNIPROT:P30406; GB:J00336; NID:G342121; PIDN:AAA36849.1; PID:G342122
 C;Superfamily: insulin
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54,90-110/Product: insulin #status predicted <MAT>
 F;55-87/Domain: insulin chain B #status predicted <BCH>
 F;90-110/Domain: insulin connecting C peptide #status predicted <CPT>
 F;31-96,43-109,95-100/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 26; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred.No. 4.4e-18;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GGGPGAGSLQPLALEGSLQKRGIVEQ 26
 |||
 Db 69 GGGPGAGSLQPLALEGSLQKRGIVEQ 94
 RESULT 4
 A42179
 Insulin precursor - chimpanzee
 C;Species: Pan troglodytes (chimpanzee)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A42179; S22058
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219553; PMID:1560757
 A;Accession: A42179

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 22, 2004, 21:32:18 ; Search time 2044 Seconds
(without alignments)
463.519 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGPGAGSLQPLALEGSLQKRGIVEQ 26

Scoring table: OLIIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65642373

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O/cgmn2_1/USPTO.spool/US10783095/runat 22112004 143631 25642/app_query.fasta_1.199
-DB=EST_QPWT-fasdbap -SUFFIX=P2Noligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=numa40.cdd -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pt
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @CGN 1 1 3437 @runat 22112004 143631 25642 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DGPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_btc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsl1:*
9: gb_gsl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26	100.0	166	4 B711381	B711381.1g94h08.x
C 2	26	100.0	168	4 BM511718	BM511718.1j52dhl1.x
C 3	26	100.0	174	5 BQ272395	BQ272395.1j98c08.x
C 4	26	100.0	179	5 BQ548672	BQ548672.1k92b08.x
C 5	26	100.0	185	4 BM857672	BM857672.1j69f01.x
C 6	26	100.0	189	5 BQ267409	BQ267409.1k07d08.x
C 7	26	100.0	190	4 BM857847	BM857847.1j69f01.x
C 8	26	100.0	190	5 B787471	B787471.1o33h10.x
C 9	26	100.0	191	5 BQ270646	BQ270646.1k06e10.x

C	10	26	100.0	191	5	BU071032	im29c07.x
C 11	26	100.0	191	5 <td>BU07517</td> <td>im83d04.x</td>	BU07517	im83d04.x	
C 12	26	100.0	191	5 <td>BU077451</td> <td>im36e11.x</td>	BU077451	im36e11.x	
C 13	26	100.0	194	4	BM565510	ih26b02.x	
C 14	26	100.0	198	5	B784487	in16a06.x	
C 15	26	100.0	203	4	BM506695	ih24b05.x	
C 16	26	100.0	220	5	BQ286797	ik34e02.x	
C 17	26	100.0	220	5	B784792	in56a07.x	
C 18	26	100.0	220	6	CA948576	ic28g01.x	
C 19	26	100.0	221	5	BQ271297	ik12a12.y	
C 20	26	100.0	221	5	BQ632376	l126c04.x	
C 21	26	100.0	225	5	BQ270528	ik04f01.x	
C 22	26	100.0	225	5	BQ272471	ij99e11.x	
C 23	26	100.0	225	5	B787442	io33e08.x	
C 24	26	100.0	226	5	B949768	in64b07.x	
C 25	26	100.0	230	5	B784626	in53h03.x	
C 26	26	100.0	231	1	A1473565	tj41a12.x	
C 27	26	100.0	232	5	B9579821	im91d10.x	
C 28	26	100.0	233	5	BQ128108	BQ128108.1j81g01.y	
C 29	26	100.0	239	5	BU074135	im84c04.x	
C 30	26	100.0	240	5	BQ270525	ik04e07.x	
C 31	26	100.0	243	5	BQ272179	ij90d09.x	
C 32	26	100.0	244	6	CA841431	iq43e09.x	
C 33	26	100.0	246	4	BM857852	l169c04.x	
C 34	26	100.0	248	6	CA842532	lq43e09.y	
C 35	26	100.0	250	6	CA847713	iq41b02.x	
C 36	26	100.0	252	5	BQ271002	ik12a12.x	
C 37	26	100.0	254	4	BM310312	ih10h08.x	
C 38	26	100.0	254	5	BQ787576	im12c06.x	
C 39	26	100.0	254	5	B948073	io49h03.x	
C 40	26	100.0	255	4	B1499525	ic58a08.y	
C 41	26	100.0	259	2	AW583461	AW583461.1a05b12.x	
C 42	26	100.0	259	2	AW583475	AW583475.1a05d09.x	
C 43	26	100.0	259	2	AW583975	AW583975.1a07h11.x	
C 44	26	100.0	261	4	B1499559	B1499559.1c60g10.y	
C 45	26	100.0	261	5	BU076352	im50d02.x	

ALIGNMENTS

RESULT 1
B711381/c
LOCUS B711381.1 GI:15687076
DEFINITION id94h08.xl Human insulinoma Homo sapiens cDNA clone IMAGE:5023718
3' similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA
sequence.
EST.
Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 166)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemshika,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theisinger,B., Ritter,E., Konko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: id94h08.y1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington

University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Seq primer: -40UP from Gibco
 Location/Qualifiers

1 .166
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5023718"
 /tissue_type="insulinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Human insulinoma"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol
 (http://genome.wustl.edu/est/lambda_protocol.shtml).
 Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:
 Pred. No.: 5,89e-15 Length: 166
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservatative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-783-095-10 (1-26) x B1711381 (1-166)

QY 1 GYGlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGlySerLeuGlnLys 20
 |||
 122 GCGGGGGCCCTGGTGCAGGCGCTGCAGCCCTGGCCCTGGAGGGTCCCTGCAGAAG 63

QY 21 ArgGlyIleValGluGln 26
 |||
 62 CGTGGCATTGTGGAACAA 45

RESULT 2
 BMS11718/c 168 bp mRNA linear EST 15-FEB-2002
 LOCUS ij52dl1.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5634500
 DEFINITION 3', similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BMS11718.1 GI:18682861
 VERSION BMS11718.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 168)
 Authors Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Elistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biochp.harvard.edu

Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Trace considered overall poor quality
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers

1 .168
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5634500"
 /tissue_type="insulinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Human insulinoma"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol
 (http://genome.wustl.edu/est/lambda_protocol.shtml).
 Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:
 Pred. No.: 6,96e-15 Length: 168
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservatative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-783-095-10 (1-26) x BMS11718 (1-168)

QY 1 GYGlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGlySerLeuGlnLys 20
 |||
 124 GCGGGGGCCCTGGTGCAGGCGCTGCAGCCCTGGCCCTGGAGGGTCCCTGCAGAAG 65

QY 21 ArgGlyIleValGluGln 26
 |||
 64 CGTGGCATTGTGGAACAA 47

RESULT 3
 BQ272395/c 174 bp mRNA linear EST 15-JUL-2003
 LOCUS ij98c08.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779574
 DEFINITION 3', similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BQ272395
 VERSION BQ272395.1 GI:20497464
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 174)
 Authors Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Elistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 22, 2004, 22:27:19 ; Search time 339 Seconds
(without alignments)
414.437 Million cell updates/sec

Title: US-10-783-095-10

Perfect score: 26
Sequence: 1 GGGGAGSLOPLALEGSLQKRGIVEQ 26

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Word size: 1

Total number of hits satisfying chosen parameters: 7244732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO_spoel/US10783095/runat_22112004_143634_25746/app_query.fasta.1.199
-DB=Published Applications_NA -OFWT=fastap -SUFFIX=P2Noligo.rnpb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095 @CGN 1.1 575 @runat_22112004_143634_25746
-NCPU=6 -ICPU=3 -NO_WMAP -LARGECQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Table with 5 columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 45 rows of sequence alignment data.

ALIGNMENTS

RESULT 1
US-10-383-285-1
Sequence 1, Application US/10383285
Publication No. US20030200566A1
GENERAL INFORMATION:
APPLICANT: STRETFIELD, STEPHEN
APPLICANT: HOWARD, JOHN
TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN PLANTS
FILE OF INVENTION:
FILE REFERENCE: 10040
CURRENT APPLICATION NUMBER: US/10/383,285
CURRENT FILING DATE: 2003-05-21
PRIOR APPLICATION NUMBER: 60/362,874
PRIOR FILING DATE: 2002-03-08
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 258
TYPE: DNA
ORGANISM: Homo sapiens
US-10-383-285-1

Alignment Scores:
 Pred. No.: 5,18e-17 Length: 258
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-10-783-095-10 (1-26) x US-10-383-285-1 (1-258)

QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGTTGGCCGGCCGGTCCCTCCAGCCGCTCGAGGCTCCCTCCAGAAG 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CCGGTATCGTGGAGCAG 210

RESULT 2
 US-09-807-742-16
 ; Sequence 16, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-807-742-16

Alignment Scores:
 Pred. No.: 5,21e-17 Length: 260
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-10 (1-26) x US-09-807-742-16 (1-260)

QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGTTGGCCGGCCGGTCCCTCCAGCCGCTCGAGGCTCCCTCCAGAAG 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CCGGTATCGTGGAGCAG 210

RESULT 3
 US-09-807-742-17
 ; Sequence 17, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 17
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
 ; OTHER INFORMATION: modified proinsulin sequence
 US-09-807-742-17

Alignment Scores:
 Pred. No.: 5,21e-17 Length: 260
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-10 (1-26) x US-09-807-742-17 (1-260)

QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGTTGGCCGGCCGGTCCCTCCAGCCGCTTAGCTTTAGAAAGSTCTTTACAAAA 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CCGGTATCGTGGAGCAG 210

RESULT 4
 US-10-746-149-23
 ; Sequence 23, Application US/10746149
 ; Publication No. US20040172667A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.
 ; APPLICANT: Fioretti, William C.
 ; APPLICANT: Cadd, Gary G.
 ; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
 ; FILE REFERENCE: 51687-0280 (51687-294923)
 ; CURRENT APPLICATION NUMBER: US/10/746,149
 ; CURRENT FILING DATE: 2003-12-24
 ; PRIOR APPLICATION NUMBER: US 60/392,415
 ; PRIOR FILING DATE: 2002-06-26
 ; PRIOR APPLICATION NUMBER: US 60/441,381
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,447
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,405
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,502
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,392
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,377
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 10/609,019
 ; PRIOR FILING DATE: 2003-06-26
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-746-149-23

Alignment Scores:
 Pred. No.: 5,21e-17 Length: 260
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n.model

Run on: November 22, 2004, 21:33:28 ; Search time 69 Seconds
(without alignments)
267.933 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGSPGAGSLOPLALEGSLQKRGIVEQ 26

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1642189

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-C=/cgn2_1/USPTO.spool/US10783095/runat_22112004_143632_25659/app_query.fasta_1.199
-DB=Issued_patents_NA -OPMT=fastap -SUFFIX=P2Noligo.rni -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095 @CGN 1 1 69 @runat_22112004_143632_25659
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOPE6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	258	2	US-08-600-783-17
2	26	100.0	269	2	US-08-673-312-10
3	26	100.0	275	1	US-07-918-953-12
4	26	100.0	275	1	US-07-918-953-14
5	26	100.0	275	1	US-08-081-661-12
6	26	100.0	275	1	US-08-081-661-14
7	26	100.0	275	1	US-07-918-953-16
8	26	100.0	276	1	US-08-081-661-16
9	26	100.0	281	1	US-07-764-655D-12
10	26	100.0	281	1	US-07-764-655D-13
11	26	100.0	281	6	5514646-1
12	26	100.0	290	1	US-08-030-731A-41

c	13	26	100.0	290	1	US-08-030-731A-42
	14	26	100.0	298	1	US-07-826-928A-28
	15	26	100.0	304	1	US-07-696-551B-12
	16	26	100.0	330	4	US-08-472-701-1
	17	26	100.0	330	5	PCT-US95-08596-1
	18	26	100.0	351	4	US-09-280-030-49
	19	26	100.0	359	3	US-08-589-028-3
	20	26	100.0	359	3	US-08-784-582-3
	21	26	100.0	359	3	US-08-785-271-3
	22	26	100.0	390	4	US-09-280-030-48
	23	26	100.0	416	3	US-08-945-140-5
	24	26	100.0	450	4	US-09-185-852-1
	25	26	100.0	450	4	US-09-323-738-1
	26	26	100.0	510	1	US-07-918-953-7
	27	26	100.0	510	1	US-08-081-661-7
	28	26	100.0	515	3	US-08-589-028-1
	29	26	100.0	515	3	US-08-784-582-1
	30	26	100.0	515	3	US-08-785-271-1
	31	26	100.0	2499	4	US-09-775-508C-7
	32	26	100.0	4992	4	US-09-015-399-10
	33	21	80.8	87	4	US-09-676-787-1
	34	21	80.8	277	1	US-07-715-183C-3
	35	21	80.8	278	1	US-07-715-183C-2
	36	19	73.1	4646	2	US-08-600-783-18
	37	19	73.1	4646	4	US-09-485-286-14
	38	17	65.4	99	1	US-07-918-953-3
	39	17	65.4	99	1	US-08-081-661-3
	40	17	65.4	598	3	US-09-537-696-11
	41	17	65.4	633	3	US-09-537-696-12
	42	17	65.4	666	3	US-09-537-696-13
	43	16	61.5	51	6	5514646-42
	44	15	57.7	74	4	US-09-485-286-9
	45	14	53.8	51	6	5514646-43

ALIGNMENTS

RESULT 1
US-08-600-783-17
; Sequence 17, Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suh
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apkujong-dong,
; CITY: Kangnam-ku
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suh
; STREET: Garden Heights Apt. 202-801, #100,

```

; STREET: Hwangkeum-dong, Soosung-ku
; CITY: Taegu
; STATE: Taegu
; COUNTRY: Republic of Korea
; ZIP: 706-040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,783
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 95-2751
; FILING DATE: 15-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-1000
; TELEFAX: (212) 953-7249
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; US-08-600-783-17

```

```

Alignment Scores:
Pred. No.: 1,9e-17 Length: 258
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-08-600-783-17 (1-258)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGlySerLeuGlnLys 20
Db 133 GCGGGTGTCCGGGTGCGAGGCTCTCCAGCGCTGGCGTGGAAAGTCCCTGCAGAAA 192
Qy 21 ArgGlyIleValGluGln 26
Db 193 CGTGGCATCGTTGAACAA 210

```

```

RESULT 2
US-08-673-312-10
; Sequence 10, Application US/08673312
; Patent No. 5891699
; GENERAL INFORMATION:
; APPLICANT: BOULAIN, JEAN-CLAUDE
; APPLICANT: CATHOLICO, LAURENCE
; APPLICANT: DUCANCEL, FREDERIC
; APPLICANT: MENEZ, ANDRE
; TITLE OF INVENTION: MODIFIED BACTERIAL ALKALINE PHOSPHATASES
; AND THEIR APPLICATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBOLON, SIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/673,312
; FILING DATE: 28-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 95-07833
; FILING DATE: 29-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBOLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 846-380-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; US-08-673-312-10

```

```

Alignment Scores:
Pred. No.: 1,97e-17 Length: 269
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

```

```

US-10-783-095-10 (1-26) x US-08-673-312-10 (1-269)
Qy 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGlySerLeuGlnLys 20
Db 138 GCGGGGGCCCGGGCGCGAGCCTCCAGCGCTGGCGTGGAGGGGCGCCCTCCAGAAG 197
Qy 21 ArgGlyIleValGluGln 26
Db 198 CGTGGCATTTGGACAG 215

```

```

RESULT 3
US-07-918-953-12
; Sequence 12, Application US/07918953
; Patent No. 5268453
; GENERAL INFORMATION:
; APPLICANT: ANDY, ROBIN J.
; APPLICANT: LARSON, ERIC R.
; TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/918,953
; FILING DATE: 19920730
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BOZICEVIC, KARL
; REGISTRATION NUMBER: 28,807
; REFERENCE/DOCKET NUMBER: 21900-2027420

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 22, 2004, 20:48:38 ; Search time 327 Seconds
(without alignments)
417.385 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGFGAGSLQPLAESLQKRGIVEQ 26

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues
Word size: 1

Total number of hits satisfying chosen parameters: 8264066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_17/USPFO_spool/US10783095/runat_22112004_143630_25620/app_query.fasta_1.199
-DB=N Geneseg 23Sep04 -OPMTS=fastap -SUFFIX=P2Noligo.rng -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCA=IGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFW=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10783095 @CGN 1 1 470 @runat 22112004 143630 25620
-NCFU=6 -ICPU=3 -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N.Geneseg 23Sep04: *
1: Genesegm1980s: *
2: Genesegm1990s: *
3: Genesegm2000s: *
4: Genesegm2001as: *
5: Genesegm2001bs: *
6: Genesegm2002as: *
7: Genesegm2002bs: *
8: Genesegm2003as: *
9: Genesegm2003bs: *
10: Genesegm2003cs: *
11: Genesegm2003ds: *
12: Genesegm2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	26	100.0	258	AAT37134	Aat37134 Proinsuli
2	26	100.0	258	ADC51568	Adc51568 Human pro
3	26	100.0	258	ADF16306	Adf16306 Human alb
4	26	100.0	258	ADH21708	Adh21708 Human lon
5	26	100.0	260	ADI04119	Adi04119 Proinsuli
6	26	100.0	261	AAT17831	Aat17831 Optimised

7	26	100.0	261	2	AAT17830	Human ins
8	26	100.0	268	2	AAQ43935	MetTyr hu
9	26	100.0	269	1	AAT80401	Proinsuli
10	26	100.0	272	1	AAAN20041	Human pro
11	26	100.0	272	1	AAAN60872	Sequence
12	26	100.0	275	2	AAQ37003	Proinsuli
13	26	100.0	279	2	AAQ83195	Proinsuli
14	26	100.0	281	2	AAQ05663	Human pro
15	26	100.0	281	2	AAQ38310	hpi gene
16	26	100.0	283	2	AAQ20543	Encodes p
17	26	100.0	289	12	ADL24436	Modified
18	26	100.0	290	12	ADL24441	Modified
19	26	100.0	298	2	AAQ27123	DNA encod
20	26	100.0	298	2	AAQ99025	Chelating
21	26	100.0	304	2	AAQ15230	Ncol-Sali
22	26	100.0	327	6	ABN85338	Oligonucl
23	26	100.0	330	9	ADA09217	Human mRN
24	26	100.0	330	12	ADJ25929	Human pre
25	26	100.0	333	3	AAZ51476	Human ins
26	26	100.0	342	1	AAAN40179	Sequence
27	26	100.0	342	1	AAAN50152	Sequence
28	26	100.0	342	1	AAAN50082	Proinsuli
29	26	100.0	351	3	AAZ59208	WMPsp-WMP
30	26	100.0	359	2	AAT75193	Human ins
31	26	100.0	359	2	AAT75652	Human ins
32	26	100.0	359	3	AAZ55717	Mutant hu
33	26	100.0	359	4	AAF58802	Human ins
34	26	100.0	390	3	AAZ59207	WMPsp-WMP
35	26	100.0	416	1	AAAN10052	mRNA sequ
36	26	100.0	416	1	AAAN40256	Sequence
37	26	100.0	450	3	AAZ29531	Human pro
38	26	100.0	450	4	AAAD17486	Human pro
39	26	100.0	450	6	ABK81193	cdNA erco
40	26	100.0	450	10	ACC78888	Humanised
41	26	100.0	450	12	ADL24432	Human ins
42	26	100.0	450	12	ADN49715	Human ins
43	26	100.0	462	3	AAZ40897	Human pre
44	26	100.0	496	3	AAA95514	Insert fr
45	26	100.0	515	3	AAZ55716	Human ins

ALIGNMENTS

RESULT 1
AAT37134
ID AAT37134 standard; DNA; 258 BP.
XX
AC AAT37134;
XX
DT 15-APR-1997 (first entry)
XX
DE Proinsulin coding sequence.
XX
KW Proinsulin; human; beta-turn peptide; insulin; A chain; protein folding;
enzymatic cleavage; B chain; miniproinsulin; enzymatic hydrolysis; ss.
XX
OS Synthetic.
XX
PN GB2298206-A.
XX
PD 28-AUG-1996.
XX
PF 14-FEB-1996; 96GB-00002998.
XX
PR 15-FEB-1995; 95KR-00002751.
XX
PA (HANI-) HANIL SYNTHETIC FIBER CO LTD.
XX
PI Shin H, Chang S, Kim D, Kim C;
XX
DR WPI; 1996-373460/38.
XX
PT New human pro-insulin derivs. - comprising A and B chains linked by small

PT beta-turn peptide.
 XX Example 4; Page 18; 52pp; English.
 XX This sequence represents a synthetic human proinsulin coding sequence.
 CC The codon preference of this sequence has been altered to favour the
 CC preferred codons in *E. coli*. This sequence can be used in the human
 CC proinsulin derivatives of the invention. In the derivatives of the
 CC invention, a beta-turn peptide (see AAW03935-W03938) is used to join the
 CC human insulin A and B chains. Beta-turn peptides are thought to be sites
 CC for initiation of protein folding. This is thought to be due to the fact
 CC that they are determined by short-range interactions, and therefore limit
 CC the conformational space available to the polypeptide chain. Beta-turns
 CC also play a valuable role in relation to enzymatic cleavage. By using the
 CC insulin derivatives (also referred to as miniproinsulin), refolding and
 CC hydrolysing processes can be carried out more efficiently than with
 CC proinsulin (which contains a bulky C-peptide). The derivatives can be
 CC recovered from transformed host cells in higher refolding yields than
 CC human proinsulin. The derivatives are also readily converted to insulin
 CC by enzymatic hydrolysis, such as with trypsin and carboxypeptidase B. The
 CC insulin derivatives are easier to produce than current derivatives. By
 CC using the derivatives of the invention, insulin can be produced at higher
 CC yields than with current techniques

XX SQ Sequence 258 BP; 50 A; 67 C; 75 G; 66 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.52e-15 Length: 258
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-10 (1-26) x AAT37134 (1-258)

QY 1 GlyGlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGGTGTCTCGGGTGCAGGGTCTCTCCAGCCGGTTCGGCTGGAGAGGTTCCCTGCAGAAA 192

QY 21 ArgGlyIleValGluGln 26
 Db 193 CGTGGCATCGTTGAACAA 210

RESULT 2
 ADCS1568
 ID ADCS1568 standard; DNA; 258 BP.
 AC ADCS1568;
 XX
 XX 18-DEC-2003 (first entry)
 DT Human proinsulin-encoding nucleoside sequence #SEQ ID 1.
 DE Antidiabetic; insulin; diabetes; genetic engineering; plant tissue; ds;
 XX human.
 KW Homo sapiens.
 OS WO2003076595-A2.
 PN 18-SEP-2003.
 XX 07-MAR-2003; 2003WO-US007316.
 PF 08-MAR-2002; 2002US-0362874P.
 PR (PROD-) PRODIGENE INC.
 XX Streetfield S, Howard JA;
 XX WPI; 2003-748386/70.
 XX

PT New plant or plant cells expressing an insulin or an insulin-like
 PT protein, useful as sources of (pro)insulin, particularly for producing
 PT (pro)insulin proteins in commercial quantities, or as oral replacement
 PT for insulin injections.
 XX Example 1; SEQ ID NO 1; 26pp; English.
 XX The invention relates to a plant, which expresses an insulin or an
 XX insulin-like protein. The plant or plant cells are useful as sources of
 CC (pro)insulin for a variety of purposes, particularly for producing
 CC (pro)insulin proteins in commercial quantities. Plant tissue can be
 CC orally administered to diabetic animals as an alternative to insulin
 CC injections. The plant-produced proteins are also useful for providing
 CC less expensive and more readily available source of the protein as
 CC reagent, or in other experimentation involving insulin proteins. The
 CC current sequence represents the human proinsulin-encoding nucleotide
 CC sequence used in an example from the invention for introduction into
 CC maize seed.

XX SQ Sequence 258 BP; 38 A; 98 C; 78 G; 44 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.52e-15 Length: 258
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-10 (1-26) x ADCS1568 (1-258)

QY 1 GlyGlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGGTGTCTCGGGTGCAGGGTCTCTCCAGCCGGTTCGGCTGGAGAGGTTCCCTGCAGAA 192

QY 21 ArgGlyIleValGluGln 26
 Db 193 CGCGGTATCGTGGAGCAG 210

RESULT 3
 ADF16306
 ID ADF16306 standard; DNA; 258 BP.
 XX ADF16306;
 AC
 XX 12-FEB-2004 (first entry)
 DT Human albumin fusion protein-related DNA sequence SeqID1398.
 DE albumin fusion protein; albumin activity; human serum albumin;
 XX serum osmotic pressure; shelf-life; stability; antidiabetic;
 KW gene therapy; diabetes mellitus; human; gene; ds.
 XX Homo sapiens.
 OS WO2003060071-A2.
 PN 24-JUL-2003.
 XX 23-DEC-2002; 2002WO-US040891.
 PF 21-DEC-2001; 2001US-0341811P.
 PR 24-JAN-2002; 2002US-0350358P.
 PR 28-JAN-2002; 2002US-0351360P.
 PR 26-FEB-2002; 2002US-0359370P.
 PR 28-FEB-2002; 2002US-0360000P.
 PR 27-MAR-2002; 2002US-0367500P.
 PR 08-APR-2002; 2002US-0370227P.
 PR 10-MAY-2002; 2002US-0378950P.
 PR 28-MAY-2002; 2002US-0382617P.
 PR 05-JUN-2002; 2002US-0383123P.
 PR 10-JUL-2002; 2002US-0385708P.
 PR 2002US-0394625P.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 22, 2004, 20:49:53 ; Search time 2542 Seconds
(without alignments)
483.687 Million cell updates/sec

Title: US-10-783-095-10
Perfect score: 26
Sequence: 1 GGGPAGSLQPLALEGSLQKRGIVBQ 26

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9044172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_epool/US10783095/runat_22112004_143630_25630/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=P2Noligo.rge -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cgi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-CUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 GCEN 1 1 3731 @runat_22112004_143630_25630 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOF=6
-FGAPEXT=7 -YGAPOF=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_rc.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 4 rows of search results.

Table with 5 columns: c, 5, 26, 100.0, 271, 6, E43909, Novel vecto. Lists various sequence identifiers and their associated scores and lengths.

ALIGNMENTS

RESULT 1
AR077833
LOCUS AR077833 258 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 17 from patent US 5962267.
ACCESSION AR077833
VERSION AR077833.1 GI:10004579
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 258)
AUTHORS Shin,H.-C., Chang,S.-G., Kim,D.-Y. and Kim,C.-S.
TITLE Proinsulin derivative and process for producing human insulin
JOURNAL Patent: US 5962267-A 17 05-OCT-1999;
FEATURES
Location/Qualifiers
source 1..258
/organism="unknown"
/mol_type="unassigned DNA"

Table with 3 columns: Alignment Scores, Pred. No., Length. Shows alignment statistics for the sequence.

US-10-783-095-10 (1-26) x AR077833 (1-258)

QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGGTGGTCCGGGTGCAGGCTCTGACCGTTGGCGCTGGAAAGTTCCCTGCAGAA 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CGTGGCATGTTGAACAA 210

RESULT 2
 A48809 261 bp DNA linear PAT 07-MAR-1997
 LOCUS A48809 Sequence 1 from Patent EP0704527.
 DEFINITION A48809
 ACCESSION A48809.1 GI:2302472
 VERSION
 KEYWORDS unidentified
 SOURCE unidentified
 ORGANISM unclassified.

REFERENCE 1
 AUTHORS Mestric,S., Punt,P.J., Valinger,R., Van and Den,H.C.
 TITLE DNA sequences encoding biosynthetic insulin precursors and process for preparation of insulin
 JOURNAL Patent: EP 0704527-A 1 03-APR-1996;
 COMMENT PLIVA PHARM & CHEM WORKS (YU)
 Other publication CN 1126761 960717
 Other publication CA 2155451 960206
 Other publication SK 97195 960207
 Other publication SI 9500250 960229
 Other publication BG 99844 960229
 Other publication CZ 9501999 960214
 Other publication PL 309882 960219.
 Location/Qualifiers

FEATURES
 source 1..261
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.14e-17 Length: 261
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-10 (1-26) x A48809 (1-261)
 QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GCGGGGCGTGGTGCAGGACCTGACCGCTTGGCGCTGGAAAGTTCCCTGCAGAA 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CGTGGCATGTTGAACAA 210

RESULT 3
 A48810 261 bp DNA linear PAT 07-MAR-1997
 LOCUS A48810 Sequence 2 from Patent EP0704527.
 DEFINITION A48810
 ACCESSION A48810.1 GI:2302473
 VERSION
 KEYWORDS unidentified
 SOURCE unidentified
 ORGANISM unclassified.

REFERENCE 1
 AUTHORS Mestric,S., Punt,P.J., Valinger,R., Van and Den,H.C.
 TITLE DNA sequences encoding biosynthetic insulin precursors and process for preparation of insulin
 JOURNAL Patent: EP 0704527-A 2 03-APR-1996;

COMMENT
 PLIVA PHARM & CHEM WORKS (YU)
 Other publication CN 1126761 960717
 Other publication CA 2155451 960206
 Other publication SK 97195 960207
 Other publication SI 9500250 960229
 Other publication BG 99844 960229
 Other publication CZ 9501999 960214
 Other publication PL 309882 960219.
 Location/Qualifiers

FEATURES
 source 1..261
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 1..258
 /note="unamed protein product; Protein sequence is in conflict with the conceptual translation"
 /codon_start=1
 /protein_id="CAA03148.1"
 /db_xref="GI:2302474"
 /translation="FVNOHLGSHVAVLYLVCGERGFYTPKTRREAEIDLQVGQVEL GSGPAGSLQPLALEGSLQKRGIVQCCCTCSLYQLENYCN"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.14e-17 Length: 261
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-10 (1-26) x A48810 (1-261)
 QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20
 Db 133 GGTGGTGGTCCGGTGCAGGACCTTCCAGCCCTTGGCGCTGGAAAGTTCCCTGCAGAA 192
 QY 21 ArgGlyIleValGluGln 26
 Db 193 CGGGCATGTTGAGCAG 210

RESULT 4
 AR069914 269 bp DNA linear PAT 18-FEB-2000
 LOCUS AR069914 Sequence 10 from patent US 5891699.
 DEFINITION AR069914
 ACCESSION AR069914
 VERSION AR069914.1 GI:7220802
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unkown.

REFERENCE 1 (bases 1 to 269)
 AUTHORS Boulain,J.-C., Cattolico,L., Ducancel,F. and Menez,A.
 TITLE Modified bacterial alkaline phosphatases and their applications
 JOURNAL Patent: US 5891699-A 10 06-APR-1999;
 FEATURES Location/Qualifiers
 source 1..269
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.17e-17 Length: 269
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-10 (1-26) x AR069914 (1-269)
 QY 1 GlyGlyProGlyAlaGlySerLeuGlnProLeuAlaLeuGluGlySerLeuGlnLys 20